

SISTEM PENILAIAN JAWABAN ESSAY OTOMATIS BERDASARKAN NILAI KEDEKATAN KALIMAT

Nur Hayatin

Teknik Infomatika, Universitas Muhammadiyah Malang / Malang

Kontak Person:

Nur Hayatin

Kantor Jurusan Teknik Informatika UMM

Jl. Raya Tlogomas 246

Malang, 65144

E-mail: noorhayatin@gmail.com

Abstrak

Penilaian (assesment) adalah salah satu hal penting dalam proses pembelajaran. Dari beberapa pilihan metode penilaian, tes tulis dalam bentuk soal uraian (essay) sering dipilih karena lebih dapat mengukur sejauh mana peserta didik memahami materi yang diajarkan. Namun tes tulis jenis ini memiliki kelemahan diantaranya butuh waktu untuk mengoreksi lembar jawaban yang ada. Hal ini dikarenakan seorang instruktur harus membaca satu per-satu pola jawaban dari peserta didik untuk disesuaikan dengan kunci jawaban. Permasalahannya, terkadang kondisi lingkungan seperti ketidak-seimbangan jumlah pengajar dan peserta didik tidak mendukung diterapkannya metode tersebut secara efektif dan efisien. Sehingga dibutuhkan solusi yang mampu mempermudah penerapan metode penilaian berupa uraian singkat dengan tidak mengesampingkan objektivitas penilaian.

Penelitian ini mengusulkan sebuah sistem penilaian jawaban essay secara otomatis. Dimana penilaian dilakukan secara komputerisasi dengan memperhitungkan nilai kedekatan kalimat atau kemiripan antara jawaban dari peserta didik terhadap kunci jawaban. Pendekatan yang digunakan adalah text mining, secara garis besar ada 3 proses utama yang dilakukan yaitu preproses, pembentukan matrik kemiripan jawaban terhadap kunci jawaban, dan pembobotan jawaban (penilaian). Untuk menghitung kemiripan jawaban terhadap kunci jawaban digunakan semantic relatedness berdasarkan Machine Readable Dictionary (MRD). Hasil pengujian menunjukkan bahwa sistem penilaian otomatis yang diusulkan memiliki akurasi sebesar 83%. Hal ini menunjukkan bahwa sistem mampu memberikan nilai yang hampir sama dengan pola penilaian yang diberikan oleh instruktur.

Kata kunci: Sistem penilaian otomatis, essay, sentence similarity, text mining

Pendahuluan

Penilaian (*assesment*) adalah salah satu hal penting dalam proses pembelajaran. Penilaian biasanya dilakukan berdasarkan 3 aspek utama yaitu ketrampilan (*skill*), pengetahuan (*knowledge*), dan sikap (*attitude*). Salah satu jenis metode penilaian adalah tes tulis. Untuk tes tulis, biasanya seorang instruktur memberikan pertanyaan seputar materi yang diujikan dalam bentuk soal pilihan ganda maupun soal uraian singkat (*essay*). Tes tulis dalam bentuk soal uraian lebih sering dipilih karena lebih dapat mengukur sejauh mana peserta didik memahami dan menjelaskan tentang materi yang telah didapatkan.

Dalam soal berupa uraian, seorang instruktur akan memberikan pertanyaan tertulis kemudian peserta didik diminta untuk menjawab soal tersebut dalam bentuk tulisan berupa uraian singkat dengan bahasa mereka sendiri. Berbeda dengan tes tulis jenis pilihan ganda, tes tulis jenis uraian akan membutuhkan waktu ekstra untuk melakukan koreksi atau penilaian jawaban. Seorang instruktur harus membaca satu per-satu pola jawaban dari peserta didik untuk kemudian disesuaikan dengan kunci jawaban yang ada. Objektivitas dan konsistensi dari instruktur sangat penting ketika harus memutuskan sebuah jawaban benar atau salah karena akan berpengaruh terhadap hasil penilaian. Permasalahannya, ada kalanya kondisi lingkungan seperti ketidak-seimbangan jumlah pengajar dan peserta didik tidak mendukung diterapkannya metode tersebut secara efektif dan efisien. Sehingga dibutuhkan sebuah solusi yang mampu mempermudah penerapan metode penilaian jawaban berupa uraian singkat dengan tidak mengesampingkan objektivitas penilaian.

Penelitian tentang penilaian jawaban otomatis telah dilakukan oleh beberapa peneliti. Penelitian [1] menggunakan metode *Latent Semantic Analysis* (LSA) untuk memberikan bobot sebuah jawaban. Metode ini menyeleksi kata penting yang terkandung dalam kalimat namun tidak memperhatikan karakteristik linguistik.

Penelitian lainnya oleh Nafik [2] yang membuat sebuah sistem penilaian otomatis dengan menghitung jarak kemiripan kalimat menggunakan algoritma *Levenshtein Distance*. Data diperoleh dari soal kuis mata kuliah Organisasi Komputer. Rata hasil pengujian antara sistem dengan penilaian manual adalah 38,50%. Sedangkan nilai terdekat adalah 4,68 antara sistem dengan penilaian manual. Algoritma *Levenshtein Distance* akan menghitung kemiripan teks berdasarkan karakter. Sehingga algoritma ini lebih cocok untuk mengoreksi ejaan, dan kurang pas jika digunakan untuk menghitung kemiripan kalimat.

Penelitian ini mengusulkan sebuah sistem penilaian jawaban uraian singkat (*essay*) secara otomatis. Dimana penilaian akan dilakukan secara komputerisasi dengan memperhitungkan nilai kedekatan teks antara jawaban peserta didik dengan kunci jawaban. Hasil dari sistem penilaian ini adalah berupa grade atau nilai kedekatan antara jawaban dengan kunci jawaban. Sebuah jawaban akan memiliki *grade* tinggi jika nilai kedekatan antara jawaban dengan kunci jawaban adalah tinggi, sebaliknya jika nilai kedekatan antara jawaban dengan kunci jawaban rendah maka *grade* yang dihasilkan juga rendah.

Perbedaan antara penelitian yang diusulkan dengan penelitian-penelitian sebelumnya adalah terletak pada metode yang digunakan untuk melakukan penilaian. Dimana dalam penelitian ini menggunakan metode pengukuran kemiripan kalimat berdasarkan semantik. Sistem penilaian jawaban otomatis menggunakan pendekatan semantik telah dilakukan oleh [3] dan [4]. Namun kedua penelitian tersebut diterapkan untuk input jawaban bahasa inggris. Dalam penelitian ini digunakan teks berbahasa indonesia sehingga metode yang akan digunakan harus disesuaikan. Jenis semantik yang digunakan pada penelitian ini adalah *semantic relatedness* berdasarkan kamus atau yang dikenal dengan *Mechine Readable Dictionary* (MRD). Dengan menggunakan MRD sistem akan mencari relasi semantik antar kata berdasarkan kamus. Dengan kata lain sebuah kata akan digali dari definisi kata yang tersimpan pada kamus. Kamus yang digunakan merupakan kamus digital Kateglo. Karena tidak hanya mengidentifikasi kata penting berdasarkan kata yang muncul pada kalimat maka metode ini diharapkan dapat meningkatkan akurasi dari output yang dihasilkan terutama ketika diimplementasikan untuk mengetahui kemiripan antara jawaban dengan kunci jawaban. Selain itu sistem penilaian otomatis ini diharapkan dapat meringankan beban instruktur yang memiliki jam kerja terbatas serta dapat meningkatkan dan menjaga objektivitas penilaian.

Metode Penelitian

Secara garis besar, metodologi yang digunakan untuk melakukan penilaian jawaban *essay* secara otomatis berdasarkan nilai kedekatan kalimat dibagi menjadi 3 proses utama yaitu *preprocessing*, pembentukan matrik kemiripan jawaban terhadap kunci jawaban, dan pembobotan jawaban (penilaian). Adapun alur dari metodologi sistem penilaian jawaban *essay* yang diusulkan dapat dilihat pada Gambar 1.

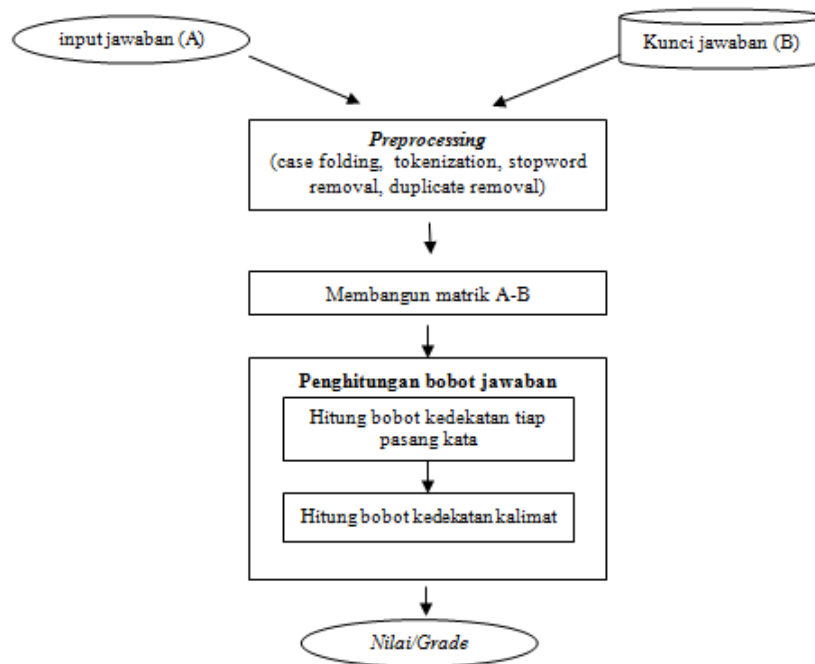
Preproses (Preprocessing)

Preproses adalah langkah awal yang dilakukan oleh sistem untuk mengolah kalimat input agar siap dilakukan proses selanjutnya. Preproses terdiri dari beberapa tahapan yaitu *case-folding*, tokenisasi, *stopword removal*, dan *duplicate removal*. *Case-folding* adalah proses normalisasi teks, dalam hal ini setiap karakter akan diubah menjadi huruf kecil atau *lowercase* (misal. “Bahasa Indonesia” menjadi “bahasa indonesia”). Selain itu karakter diluar huruf akan dihapus. Selanjutnya kalimat hasil dari *case-folding* akan dipecah kata per-kata melalui tahapan tokenisasi. Pada proses tokenisasi akan dilakukan pemenggalan atau segmentasi kata berdasarkan spasi dimana setiap kata hasil pemenggalan disebut dengan *term* (misal. “bahasa indonesia”, setelah proses tokenisasi akan dihasilkan 2 *term* yaitu “bahasa” dan “indonesia”). Tahapan berikutnya yang dilakukan adalah *stopword removal*. *Stopword* dapat diartikan sebagai kata yang tidak memiliki pengaruh yang berarti terhadap makna yang terkandung dalam sebuah kalimat, contohnya adalah kata dari, yang, untuk. *Stopword removal* adalah proses penghapusan *stopword* yang bertujuan untuk mengekstraksi kata penting (*keyword*). Proses ini selain untuk menghapus kata yang tidak memiliki pengaruh penting juga untuk mereduksi dimensi kata. Tahapan terakhir yang dilakukan pada preproses adalah *duplicate*

removal. Proses ini akan menghapus kata duplikat yang ada pada kalimat. Sehingga nantinya hasil dari proses ini adalah sebuah kalimat yang memiliki kata unik atau tidak ada kata yang sama.

Pembangunan matrik

Proses utama yang dilakukan selanjutnya adalah pembentukan matrik kemiripan jawaban terhadap kunci jawaban. Disini yang akan terbentuk kita sebut saja dengan matrik $A-B$, dimana A merupakan *term* yang diambil dari jawaban sedangkan B merupakan *term* yang diambil dari kunci jawaban. Baris pada matrik berisi kata penting hasil ekstraksi $A (A_1, \dots, A_i)$. Sedangkan kolom matrik berisi kata penting hasil ekstraksi $B (B_1, \dots, B_j)$.



Gambar 1. Metodologi Sistem Penilaian Jawaban Essay Otomatis berdasarkan Nilai Kedekatan Kalimat.

Penghitungan Bobot Jawaban

Bobot dari matrik $A-B (X_{ij})$ adalah nilai kedekatan kata antara A_i dan B_j yang dihitung berdasarkan MRD. Setelah tiap pasangan kata atau *term* diketahui bobotnya, selanjutnya dihitung kemiripan atau kedekatan antara dua kalimat dengan menggunakan metode *cosine similarity*, lihat persamaan 1. Hasil perhitungan dari persamaan 1 inilah akan menjadi bobot jawaban.

$$similarity(\vec{a_i}, \vec{a_j}) = \frac{\vec{a_i} \cdot \vec{a_j}}{|\vec{a_i}| \cdot |\vec{a_j}|} = \frac{\sum_{k=1}^n (w_{ki} \cdot w_{kj})}{\sqrt{\sum_{k=1}^n w_{ki}^2 \cdot \sum_{k=1}^n w_{kj}^2}} \quad (1)$$

Data sampel

Data yang digunakan dalam penelitian ini menggunakan sampel beberapa jawaban uraian singkat beserta kunci jawaban untuk setiap pertanyaan. Data diambil dari data kelas E-bisnis Teknik Informatika Universitas Muhammadiyah Malang. Jumlah mahasiswa dalam satu kelas ada 30 mahasiswa. Soal diambil dari soal UAS (Ujian Akhir Semester) Tahun akademik 2013/2014. Jumlah soal ada 3 soal yang berupa uraian.

Hasil Penelitian dan Pembahasan

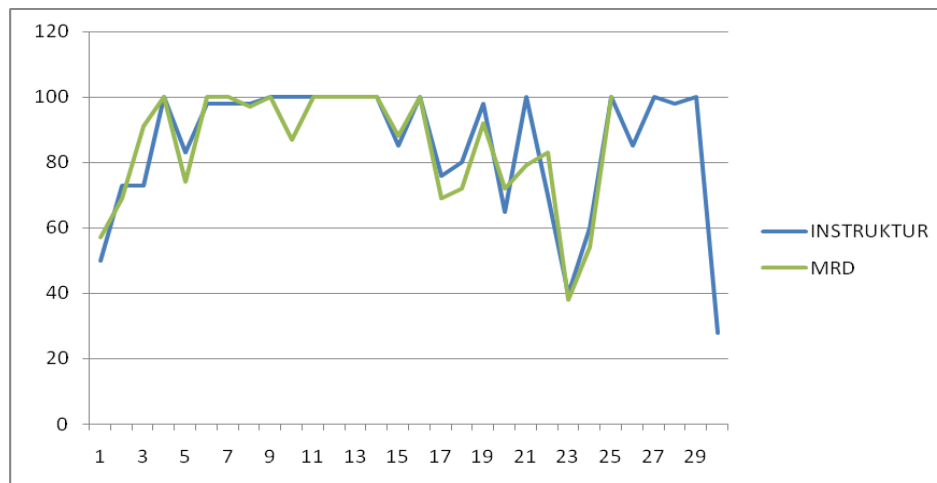
Langkah pengujian sistem penilaian otomatis yang pertama dilakukan adalah menghitung total nilai (*grade*) untuk tiap mahasiswa. Membandingkan nilai total yang dihasilkan sistem dengan nilai yang diberikan instruktur dan menghitung berapa akurasi. Tabel 1 berisi sampel dari 5 data perbandingan penilaian jawaban yang diberikan oleh instruktur terhadap sistem.

Tabel 1. Sampel perbandingan penilaian jawaban oleh instruktur terhadap sistem

NILAI INSTRUKTUR	SISTEM
50	57

73	69
73	91
100	100
83	74

Gambar 2 menunjukkan grafik dari hasil perbandingan antara penilaian yang diberikan oleh instruktur (garis warna biru) terhadap sistem (garis warna hijau) untuk 30 data. Dari grafik tersebut dapat dilihat bahwa pola dari garis hijau mengikuti pola dari garis biru dibanding dengan garis merah. Hal ini menunjukkan bahwa hasil penilaian oleh sistem hampir mendekati pola penilaian yang diberikan oleh instruktur. Dari hasil perbandingan penilaian yang diberikan oleh instruktur terhadap sistem didapatkan nilai akurasi sebesar 83%. Hal ini membuktikan bahwa sistem penilaian jawaban essay otomatis yang diusulkan hampir mendekati pola penilaian yang diberikan oleh instruktur.



Gambar 2. Grafik perbandingan nilai instruktur dan hasil generate sistem

Kesimpulan

Sistem penilaian jawaban essay otomatis yang diusulkan mampu memberikan akurasi sebesar 83%. Hal ini membuktikan bahwa sistem tersebut hampir mendekati pola penilaian yang telah diberikan oleh instruktur.

Referensi

- [1] Harisma NZ. Implementasi Sistem Penilaian Esai Otomatis Metode LSA dengan Tiga obot Kata Kunci. Laporan Skripsi. Jakarta : Fakultas Elektro Universitas Indonesia; 2008.
- [2] Nafik M.Z. Sistem Penilaian Otomatis Jawaban Esai menggunakan Algoritma Levenshtein Distance. Universitas Brawijaya; 2012.
- [3] Firdausiah, A.B. Sistem penilaian Otomatis Jawaban Essay menggunakan Ontologi pada Moodle. Telkomnika. 2008; 6 (3).
- [4] Winarsono D. Sistem Penilaian Otomatis Kemiripan Kalimat menggunakan Syntactic-Semantic Similarity pada sistem E-learning. Kursor 2009; 5(2).
- [5] Liu Hui. Computing Semantic Similarities based on Mechine-Readable Dictionary. IEEE International Workshop on Semantic Computing and Systems. 2008